

A Method to Reduce the Modifiable Areal Unit Problem in Measuring Urban Form Metrics

Dr. Abdul Naser Arafat

Keywords: urban, modifiable Areal Unit Problem, Transportation planning, GIS

ABSTRACT

The relationship between urban form and transportation is identified as being bi-directional where the urban form affects transportation and transportation affects the urban form. To capture the impact of urban form on transportation, land-use descriptors such as diversity, density, design, destinations and distance are traditionally used. Although the new trend in transportation planning is to perform disaggregate analysis in time and space, many of these urban-form characteristics are measured on an aggregate scale. However, the characteristics of a predefined areal unit affect the values of an urban form measurement. This problem is known as the Modifiable Areal Unit Problem (MAUP).

The MAUP literature explains the effect of location, size and shape of the areal units on the aggregated values captured by them. The focus of this research is on the use of an optimized zoning procedure to reduce the effect of the MAUP on the estimation of urban form patterns. To do that, an extensive parcel level GIS iterative analysis will be performed on three urban counties in the state of Florida in the United States.

To minimize the zonal boundary impact, every parcel will be taken as a center of a zone that will capture the urban form characteristic surrounding that parcel, and then the analysis will be repeated for different zone sizes and shapes to find an optimal size for the zones. The research will show that the

use of natural neighborhood boundaries as an areal unit may lead to undesirable results because of the differences in size and shape. The research will also propose a GIS tool that uses the optimal neighborhood in measuring urban form characteristics. Diversity (land use mix), design (connectivity), destination (accessibility) and distance (5Ds) are used to study the interaction between land use

and transportation, especially with regard to the impact of land use on transportation in terms of vehicle trips and vehicle miles of travel (VMT) (Ewing and Cervero, 2001; Ewing et al, 2008; Lee & Cervero, 2007). The five Ds in general represent some of the measurements for the urban form characteristics. Therefore, their impact is not isolated from scale and zoning problems.

Urban-form characteristics such as density, diversity and connectivity are usually measured on an aggregate scale (e.g. see measuring land use patterns; Hess et al, 2003; Dill, 2004). However, the scale and zoning of the predefined areal unit affect the values of the urban form measurement. This problem is known as the Modifiable Areal Unit Problem (MAUP) (Guo & Bhat, 2004; Kwan & Weber, 2008).

In the MAUP, the scale and zoning of the areal unit affect the value and variation of the aggregated entity. The MAUP is a recognized problem that has faced researchers in the modeling and calculation of entities that have a spatial representation. The scale

and zoning in this research are represented by the areal unit characteristics which include but are not limited to the size, shape and location of the areal unit.

Geographic information system GIS can be used to reduce the scale and zoning issues and offers solutions to reduce the modifiable areal unit problem; GIS Spatial Analyst offers the tools for defining the areal unit and running iterative analysis on different sizes and shapes. This paper will present the impact of size and shape of the areal unit on the urban form characteristics and will present a method to optimize the neighborhood size and shape. For the location of the areal unit, the paper will present a floating areal unit method to reduce the effect of the zoning and adjacent neighborhoods limitation.

Literature Review:

Urban form metrics such as density, diversity of land uses, connectivity are usually estimated after defining the areal unit of analysis such as the neighborhood or any type of gridding. However, not all the urban form characteristics should be aggregated to a defined areal unit. Sometimes it is better to understand the values on a more disaggregate level such as the parcel level. In this paper, accessibility and distance are not included in the MAUP research, because these variables are recommended to be performed on a disaggregate level by many researchers. Understanding what is accessibility helps to understand the recommendation for disaggregate analysis for accessibility.

The measurements of accessibility vary from linear distances, network distances, travel times or the number of activities within a distance from an attraction or a certain residential location. Many mathematical forms are used to estimate accessibility. Bhat et al (2002), summarized accessibility measurements into different equations for cumulative opportunity and gravity. These estimations are classified as Gaussian, composite impedance, activity distance and in-vehicle travel time.

Common methods of capturing accessibility are based on aggregate analysis zones such as Traffic Analysis Zones (TAZ) (For example, see accessibility estimation, Levinson & Krizek, 2008). However, it is

more helpful and more accurate to go to the parcel level. Disaggregate and parcel level research can reduce the shortcomings of the traditional models in capturing the fine land-use effect on transportation or vice versa (Lee, 2004). Primarily there are two approaches in the levels of analysis used to delineate or determine the connection between land-use and transportation.

One of them is an aggregate approach that uses zones such as traffic analysis zones (TAZs) and calculates accessibility indices based on the analysis zones. The other uses parcel-level analysis. Johnston (2004) mentioned that future land-use and transportation modeling should be discrete in both time and location and based on GIS tract or street address. This shows the common ground for future land-use and transportation modeling. The same GIS approach had been recommended by Wegner (2005) to deal with the disaggregate data in activity-based models, which is fundamental in the latest trend in transportation modeling. However, discrete analysis may not always be possible due to lack of data.

Therefore, TAZ and neighborhood urban form aggregation is used to conduct land-use and transportation analysis. However, if TAZs are used, the effect of the MAUP should be taken into account for calculating the aggregate accessibility value as well as its spatial variation. In practice, most of the mathematical forms for accessibility are used on aggregate levels such as TAZ. The same equations can be applied on a parcel-level to give more accurate accessibility measurements and eliminate the effect of the MAUP for accessibility estimation. Sometimes accessibility is applied in an equation on a parcel-level and the resultant accessibility index is aggregated to a neighborhood level. In this case the areal unit problem should also be taken into account.

Several land use transportation coordination variables are urban form characteristics that are aggregate measures by definition, and thus measured on a neighborhood scale or other defined analysis unit areas. The neighborhood scale primarily will affect the value of measurement. Ewing et al (2008) found the impact of the 5Ds on VMT is less for small areas, thus the elasticity values should differ. The differences between values of land use and transportation characteristics aggregated to

the neighborhood level are caused by the MAUP. According to Jelinski and Wu (1996) the MAUP has two problem components which are the scale and the zoning problems. In the scale problem, the value and the statistical variation of the aggregated smaller units will be different depending on the size of the areal unit. However, the zoning problem deals with location of the areal unit and the choice of zoning which also affect the value and the variation of the aggregated entities. Due to the zoning problem, the same size aggregated units may give different results due to differences in location. Researchers frequently suggest grid based approaches to do the spatial aggregation. However, little attention is paid to the size of the grid cells and the method of aggregation to reduce the MAUP.

Reynolds (1998) mentioned the MUAP as a problem in spatial research as early as the 1930s but indicated it could be reduced by the use of GIS. He also mentioned that many of the data are collected on a disaggregate level and aggregated to coarser resolution because of different reasons such as privacy. However, most of the time data is measured at a disaggregate level but mapped spatially on an aggregate level. Therefore, the aggregation process will be necessary to map the data. Reynolds also explained the two sides of the areal unit problem which are the scale and the zoning.

Tomoki (1999) explained the dependence of the map interpretation on the areal unit and emphasized that currently there are only a few criteria on the choice on the areal units. Tomoki explained the effect of the areal unit size on mortality indicators where the variation in the value decreases for larger units. Therefore taking larger units will be more statistically stable but at the same time the results will be more ambiguous. In summary, the size, shape and boundaries of neighborhoods affect the value of any aggregate measurement based on these neighborhoods.

General gridding procedures and naturally defined neighborhoods are used in the research to define unit areas. If one is to research the land use mix variations using an entropy value on a parcel adjacent to the boundary in a naturally defined neighborhood; this parcel may have, for example, retail and service opportunities surrounding it and the entropy value for that neighborhood will not

capture that land use mix. This is mainly a form of the zoning effect induced by the MAUP.

Steiner & Srinivasan (2009) proposed a solution for this problem by using overlapping 2 by 2 mile square neighborhoods for calculating the land use and transportation characteristics for their trip length model. To reduce the edge effect of the MAUP, a methodology to overlap neighborhoods is used where the neighborhoods were duplicated and shifted 1 mile east and the resultant duplicated neighborhood are duplicated again and shifted 1 mile north. Therefore, a parcel at the edge of a neighborhood will be at the center of an overlapping neighborhood.

Taking into consideration the scale and zoning components of the MAUP, it is clear that the size, location and shape of the areal unit are the main component that may affect the value and the variation of the urban form measurements aggregated to that area. Many of the urban form measurements such as accessibility are recommended to be performed on a parcel level and therefore no need to consider the modifiable area problem. However, other urban form measurements such as density, land use mix and connectivity are needed on an aggregate scale. Therefore, the effect of the modifiable areal unit should be taken into account when dealing with those variables.

Methodology:

A first step to obtain the optimal neighborhood for capturing land use and transportation characteristics is researching the effect of size and shape as well as the boundaries of neighborhoods used for the data aggregation. To do that, preliminary research is performed using variable neighborhood size around each parcel to study the effect of neighborhood size and shape on the land use mix entropy value. The same method is applied on other urban form characteristics such as connectivity and density. The methodology accounts for the scale component of the MAUP by taking a variable size neighborhood. The methodology also accounts for the zoning component by using a floating areal unit as well as different unit shapes.

The land use mix in this research is captured by the entropy measurement as defined by Frank & Pivo (1994). The other two variables studied here are the connectivity and density. For neighborhood

connectivity, a simple measurement of road density is used as a connectivity indicator.

This measurement is primarily miles of roads per unit of area (square miles). Other connectivity measurements can be used too, such as the number of nodes per links in the neighborhood (Dill, 2004). ArcMap's Spatial Analyst extension is used to create a raster representing the connectivity for a certain defined areal unit. For the density, this research uses people per acre variable from the Census block data. These density values are aggregated and averaged to the defined areal unit. The value of the average density is assigned to the center point of the areal unit.

To eliminate the zoning boundary problem, a floating areal unit is used where each point in the map is a center of a defined neighborhood. In summary, a multi- level analysis for land use mix, density and connectivity will be performed using a neighborhood size that will be increased incrementally according to the following conditions:

1. Because of the zoning component of the MAUP, each parcel will be taken as a center of this neighborhood and the analysis will be replicated for all parcels.
2. The result surfaces will be compared by ArcGIS statistics by calculating the mean and standard deviation for the change in value and standard deviation due to the incremental change in size.

The following figure shows the scale component by displaying the variable size unit area around the parcel.

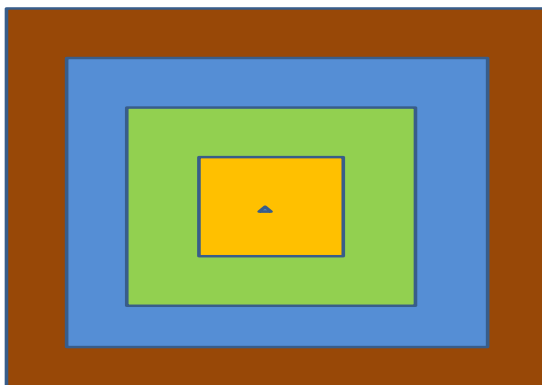


Figure 1: Variable Size Neighborhoods around the Parcel. The smallest size is a 0.5 mile and the increment is taken as .5 mile

To study the effect of the areal unit shape, three different unit shapes are studied. These shapes are a square, a circle and a diamond. The square shape represents aggregated parcel blocks in a grid pattern system. The circle represents the Euclidian travel distance, and the diamond shape represents the Manhattan travel distance. The results of the three shapes are overlaid with the street network to obtain the optimal areal unit shape.

ArcGIS Model Builder models and customized Python scripts are used to calculate and create surfaces for the land use mix (entropy), density and connectivity using floating areal units. Multiple shapes and sizes are used in the analysis. The models and scripts calculate the value for each parcel assuming that it is at the center of the areal unit. After generating the surfaces, statistical indicators using raster statistics are generated and used to obtain the optimal neighborhood size and shape.

Results and Discussion:

This paper focuses on the optimization of the areal unit specifically for the research on land use and transportation coordination. Many of the urban form indices can be calculated on disaggregate level such as parcel-level analysis but the results are usually aggregated to a neighborhood or any other aggregate scale areal unit such as TAZs. This research studies the effect of the scale and zoning of the areal unit by investigating the impact of size, shape and location of the areal unit on the most commonly used urban form measurements for land use and transportation coordination.

These urban form measurements are diversity, density and connectivity. The unit area optimization is conducted by calculating the land use mix index represented by entropy values based on different areal unit sizes. The same procedure is applied for density and connectivity. Generally, the analysis follows an iterative procedure taking unit sizes of (0.5 x 0.5 mile - 5x5 mile) using a 0.5 mile incremental increase. The optimal neighborhood is taken as the minimum size where increasing the neighborhood no longer has a significant effect on the entropy value. Table 1 and figures 2a and 2b shows the change in mean and standard deviation due to the increase of areal unit size for estimating the entropy in three counties in Florida.

| Size From | Size To | Mean Change Orange | Mean Change Duval | Mean Change Hills-borough | STD Change Orange | STD Change Duval | STD Change Hills-borough |
|-----------|---------|--------------------|-------------------|---------------------------|-------------------|------------------|--------------------------|
| 0.5 | 1 | 0.0784 | 0.0484 | -0.0309 | 0.0911 | 0.0832 | 0.0685 |
| 1 | 1.5 | 0.0199 | 0.0267 | 0.0006 | 0.0484 | 0.0558 | 0.0387 |
| 1.5 | 2 | -0.1126 | 0.0177 | 0.0051 | 0.0924 | 0.0406 | 0.0281 |
| 2 | 2.5 | 0.0038 | 0.0126 | 0.0036 | 0.0204 | 0.0315 | 0.0221 |
| 2.5 | 3 | 0.0035 | 0.0097 | 0.0028 | 0.0165 | 0.0257 | 0.0181 |
| 3 | 3.5 | 0.0031 | 0.0077 | 0.0027 | 0.0139 | 0.0213 | 0.015 |
| 3.5 | 4 | 0.0028 | 0.0064 | 0.0022 | 0.0124 | 0.0182 | 0.0112 |
| 4 | 4.5 | 0.0026 | 0.0054 | 0.0032 | 0.011 | 0.0158 | 0.0132 |
| 4.5 | 5.0 | 0.0025 | 0.0046 | 0.0023 | 0.0097 | 0.0138 | 0.0102 |

Table 1: Change in Entropy Mean and Standard Deviation for Three Counties

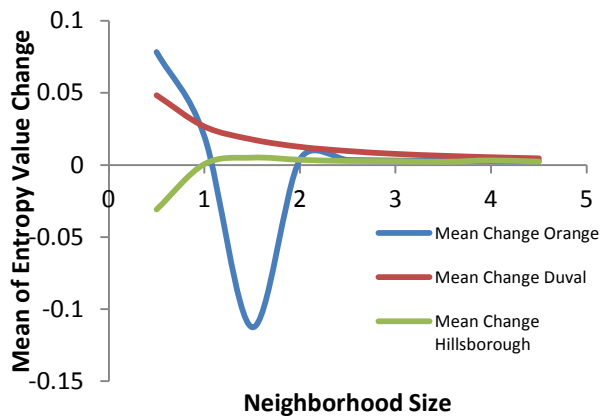


Figure 2a: Mean of entropy value change that corresponds to 0.5 mile change in unit size

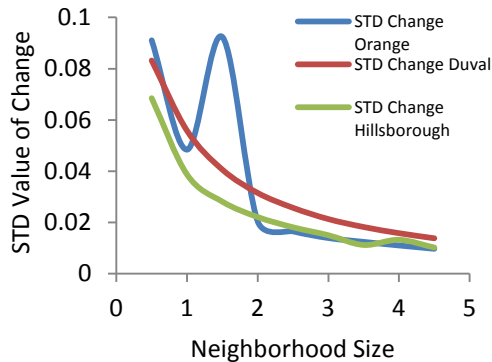


Figure 2b: Standard deviation of entropy value change that corresponds to 0.5 mile change in unit size

The result shows also that the variation in entropy values are less for larger neighborhoods which agrees with the literature on the scale component of the MAUP. The following figure shows the entropy surface based on a 2.5 x 2.5 mile neighborhood for Orange County. Conducting the analysis on a roving unit where the calculated cell value is always in the

center of the neighborhood minimizes the zoning problem and creates a surface of a finer resolution.

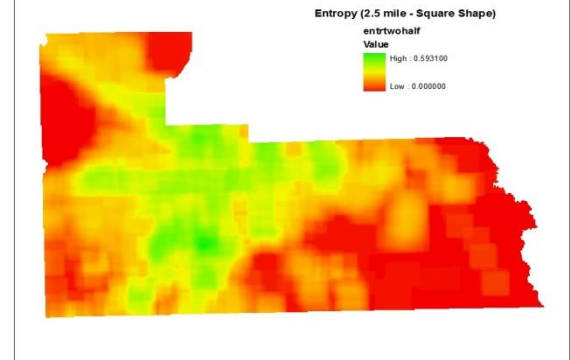


Figure 3: Land use mix entropy based on 2.5x2.5 mile areal unit for Orange County

Density is usually calculated on multiple scales depending on it is use. However, in land use and transportation research the neighborhood density is used to identify the impact of density on travel cost or VMT.

In this paper, a simple density measurement will be used which is the gross population density per acre averaged for each the areal unit and assigned to the point at the center of the neighborhood.

Table 2 and figures 4a and 4b shows the change in mean and standard deviation for density when increasing the areal unit size for orange. Similar results were obtained for Duval and Hillsborough counties.

| Size From | Size To | Mean Change Orange | STD Change Orange |
|-----------|---------|--------------------|-------------------|
| 0.5 | 1 | -0.337 | 1.286 |
| 1 | 1.5 | 0.012 | 0.502 |
| 1.5 | 2 | 0.003 | 0.349 |
| 2 | 2.5 | 0.002 | 0.259 |
| 2.5 | 3 | 0.001 | 0.203 |
| 3 | 3.5 | 0.001 | 0.168 |
| 3.5 | 4 | 0.001 | 0.145 |
| 4 | 4.5 | 0.001 | 0.130 |
| 4.5 | 5.0 | 0.001 | 0.116 |

Table 2: Change in density mean and standard deviation

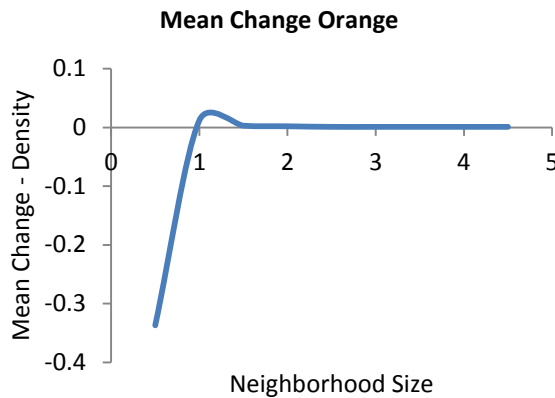


Figure 4a: Mean of density value change that corresponds to 0.5 mile change in unit size



Figure 4b: Standard deviation of density value change that corresponds to 0.5 mile change in unit size

The result from the three counties was that the neighborhood size of 2.5 mile by 2.5 mile can be used as an optimized neighborhood size for density calculation for the three counties. The following figure shows the density surface for Hillsborough County based on 2.5 mile unit.

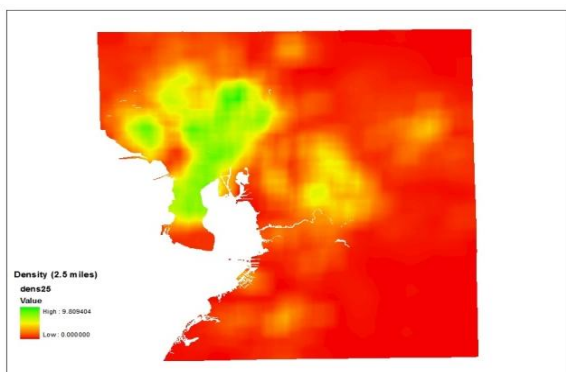


Figure 5: Gross population density surface for Hillsborough County based on 2.5x2.5 mile areal unit

For the connectivity measurement, the street density is another urban form measurement to be tested. The following table shows the mean value and the standard deviation of the change in road density values resulted by a gradual increase of 0.5 miles in units size. Table 3 and figures 6a and 6b show that the minimal change in road density mean value corresponds to a unit size of 2.5 x 2.5 miles or more

| Size From | Size To | Mean Change Duval | STD Change Duval |
|-----------|---------|-------------------|------------------|
| 0.5 | 1 | -4.437 | 3.529 |
| 1 | 1.5 | -4.384 | 1.197 |
| 1.5 | 2 | -0.043 | 1.351 |
| 2 | 2.5 | -0.043 | 1.027 |
| 2.5 | 3 | -0.042 | 0.82 |
| 3 | 3.5 | -0.041 | 0.671 |
| 3.5 | 4 | -0.04 | 0.572 |
| 4 | 4.5 | -0.039 | 0.491 |
| 4.5 | 5.0 | -0.038 | 0.429 |

Table 3: Change in connectivity mean and standard deviation

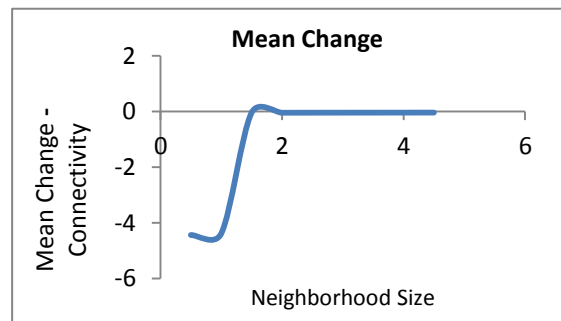


Figure 6a: Mean of connectivity value change that corresponds to 0.5 mile change in unit size

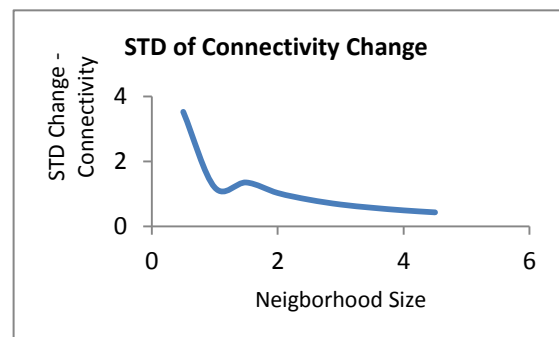


Figure 6b: Standard deviation of connectivity value change that corresponds to 0.5 mile change in unit size

The analysis is replicated on three urban counties and it was found that the neighborhood size of 2.5 mile by 2.5 mile can be used as an optimized neighborhood size for connectivity calculation for the three counties. The following figure shows the connectivity surface for Duval County based on a 2.5 mile areal unit.

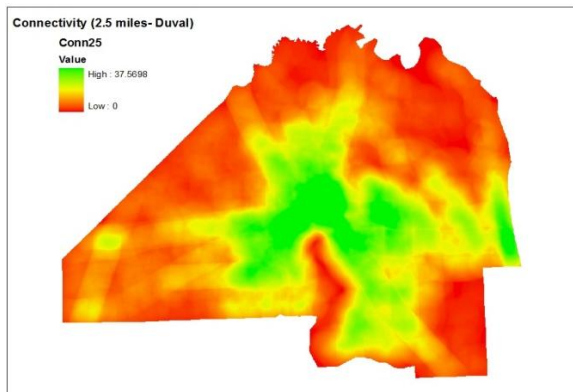


Figure 7: Connectivity surface for Duval County based on a 2.5 x 2.5 mile areal unit

Additionally, the optimal neighborhood size from the three counties is tested for the change in the shape. A circle and diamond shaped boundary instead of a square shaped boundary for the neighborhood analysis are used to test the impact of the neighborhood shape on the entropy. Figure 8 shows the different areal unit shapes used in the analysis.

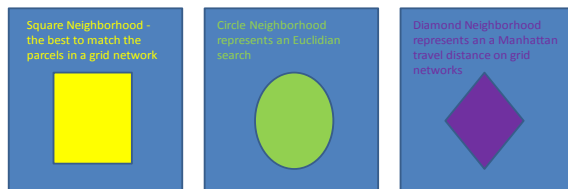


Figure 8: Shapes for the Tested Areal Units

Figure 9 is the output entropy surface for Hillsborough using a 2.5 mile of a diamond shaped unit which represents a driving / biking distance of 1.25 miles. To understand the results of changing the neighborhood shape, the entropy layer is overlaid with other layers such as street networks as shown in Figure 10. It was found that the land use mix layer resulting from the shape analysis using the Manhattan diamond shape is more subjectively justified than the result surfaces using square and circle neighborhoods. This is because the diamond shaped neighborhood gives more mixed use density

in areas that have more street networks while other surface have not distinguish these areas.

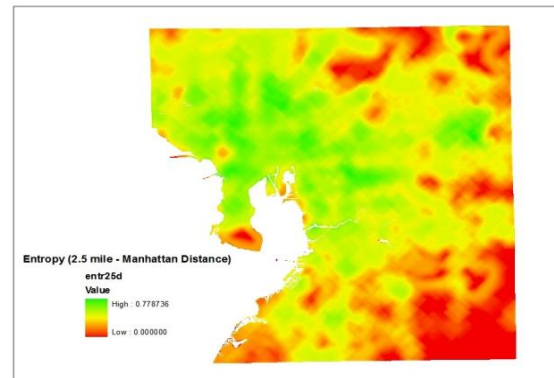


Figure 9: Land mix entropy surface for Hillsborough County based on 2.5 mile diamond shaped areal unit



Figure 10: Streets and land use mix entropy overlay surface for Hillsborough County based on 2.5 mile diamond shaped areal unit.

CONCLUSION, LIMITATIONS AND FUTURE RESEARCH

The analysis in this research proved that changing the areal unit size of a neighborhood changes the aggregated value of the variable being measured. The generated surfaces also show that the variation in aggregated value increases for smaller zones and decreases for larger zones. Therefore the results show that larger areal units are more stable but the values could be more ambiguous which is corroborated by the literature on the MAUP.

The results also show that the neighborhood size could be optimized for the tested urban form variables which are land use mix, density and connectivity. The graphs for the change in aggregated values corresponds to the change in neighborhood size and shows that the curve is nearly

flat for neighborhoods sized 2.5 mile and more, and the standard deviations for the change are very low. This means that the optimal size of the neighborhood is the minimum unit size where the curves began to be flat which is 2.5 mile x 2.5 mile. For the scale issue of the MAUP, the results also show that the change in value is very large for small neighborhood sizes, which means that using small unit sizes as area units may lead to undesirable results. On the local level, the entropy value change is large and unpredictable as shown in figure 3.

Therefore changing from a 0.5 mile to a 1 mile size could positively or negatively change the aggregated values. It is true that the variation is less for larger neighborhood, but from the results, it is not difficult to conclude that the change in aggregated values due to changing neighborhood sizes in small neighborhood is not predictable. However, this will raise the question of how applicable the natural neighborhood will be as an area unit where the unit sizes are always different between zones.

From the results of the analysis we can conclude that it is better to use the same size of area unit for the whole spatial environment and changing the unit size may lead to undesirable results. Taking entropy as an example, the magnitude and direction of change in entropy values for small neighborhoods gives us a hint that the use of an entropy value, in a naturally defined neighborhood may lead to unpredictable results due to the MAUP. This may be reduced if the same neighborhood size for the entire model is used. The problem can be also reduced if the optimized neighborhood size is used as well as the floating area methodology.

For the zoning issue in the MAUP, the results show that changing the boundary shape leads to different aggregated results, which confirm the MAUP zoning problem. The results show that the shape can be optimized by analyzing the urban form and the spatial distribution of activities. However, analyzing the land use mix relationship with the street network, we can see that the diamond shape which represents a Manhattan grid driving or biking distance gives better aggregate results than the Euclidean distance (circle) or the square neighborhood.

The zoning problem in this research is also reduced by the use of the floating neighborhood where each

point in the map is taken as the center of the aggregated zone. However, this unit area definition works for the purpose of tested variables in this research and may not be valid for other types of neighborhood aggregation processes.

The research in this paper has many limitations that can be summarized as follows: Firstly, the research is limited to the urban context of the three tested counties. The counties may have different spatial patterns but the three counties have large cities and CBD's. The research is not applied to rural counties and that will be left for future research.

Secondly, the change in aggregated value is compared by the mean and the standard deviation of change. More statistical indicators could be used in the optimization process. This also will be performed in future research.

Thirdly, the research is limited to the tested variables for land use mix, density and connectivity. Therefore the results should not be generalized.

Finally, the research is limited to the aggregation of urban form measurements that can be used to test the impact of urban form on VMT or trip length. The optimized unit should not be used for the aggregation of values that depend on small neighborhood sizes by definition such as walking or transit zones.

References:

- Ewing, R., & Cervero, R. (2001). Travel and the built environment: A synthesis. *Transportation Research Record: Journal of the Transportation Research Board*, 1780(-1), 87-114.
- Frank, L. and G. Pivo (1994). Impacts of Mixed Use and Density on Utilization of Three Modes of Travel: Single-Occupant Vehicle, Transit, and Walking. *Transportation Research Record*, 1466(-1), Washington, D.C.
- Johnston, R. A. (2004). The urban transportation planning process. In Hanson, Susan and Giuliano, Genevieve (Ed.), *The Geography of Urban Transportation* (3rd ed., pp. 115-140). New York: The Guildford Press.

- Lee, B. (2004). Parcel-level Measure of Public Transit Accessibility to Destinations, Proceeding of the Access to Destination Conference, November 8–9, 2004 Minneapolis, Minnesota.
- Steiner, R.L., & Srinivasan, S. (2009). Vehicles miles of travel based traffic impact assessment empirical analysis. Unpublished manuscript.
- Guo, J.Y. and Bhat, C.R., 2004. Modifiable areal units: a problem or matter of perception in the context of residential location choice modeling? *Transportation Research Record*, 1898, 138-147.
- Hess P. Moudon. A and Logsdon M.(2003). Measuring Land Use Patterns for Transportation Research. *Transportation Research Record* 1780, 01-2512.
- Kwan M and Weber J (2008). Scale and accessibility: Implications for the analysis of land use-travel interaction. *Applied Geography*, 28(2): 110-123.
- Bhat C.R., S. Handy, K. Kockelman, H. Mahmassani, A. Gopal, I. Srour, L. Weston.(2002)"Development of an Urban Accessibility Index. Formulations, Aggregation and Application". University of Texas at Austin. Report No. FHWA/TX-02-4938-4. October, 2002.
- Dill J (2004). Measuring Network Connectivity for Bicycling and Walking. *83rd annual meeting of the transportation research board*. Washington DC.
- Ewing, R.H., Bartholomew, K., Winkelman, S., Walters, J. & Chen, D. (Eds.). (2008). *Growing cooler : The evidence on urban development and climate change*. Washington, D.C.: ULI.
- Lee, R.W., and Cervero, R. (2007). *The effect of housing near transit stations on vehicle trip rates and transit trip generation: A summary review of available evidence*. Retrieved February 18, 2010 from
- Jelinski, D. and Wu, J. (1996). The modifiable areal unit problem and implications for landscape ecology *Landscape Ecology*. *Earth and Environmental Science*, 0921-297. Springer Netherlands.
- Tomoki N. (1999). An information statistical approach to the modifiable areal unit problem in incidence rate maps. Department of Geography, Ritsumeikan University, Kita-ku, Kyoto 603-8577, Japan.
- Reynolds H. (1998). The Modifiable Area Unit Problem: Empirical Analysis by Statistical Simulation. Graduate Department of Geography. University of Toronto